

<sup>1</sup> Fitness tracking reveals task-specific associations  
<sup>2</sup> between memory, mental health, and physical activity

<sup>3</sup> Jeremy R. Manning<sup>1, \*</sup>, Gina M. Notaro<sup>1</sup>, Esme Chen<sup>1</sup>, and Paxton C. Fitzpatrick<sup>1</sup>

<sup>4</sup> <sup>1</sup>Dartmouth College, Hanover, NH

<sup>5</sup> \*Address correspondence to jeremy.r.manning@dartmouth.edu

<sup>6</sup> July 16, 2022

7

## Abstract

8 Physical exercise activity can benefit both physical and mental well-being. Different forms  
 9 of exercise (e.g., aerobic versus anaerobic; running versus walking, swimming, or yoga; high-  
 10 intensity interval training versus endurance workouts; etc.) impact physical fitness in different  
 11 ways. For example, running may substantially impact leg and heart strength but only moderately  
 12 impact arm strength. We hypothesized that the mental benefits of exercise physical activity  
 13 might be similarly differentiated. We focused specifically on how different forms-of-exercise  
 14 intensities-of-physical-activity might relate to different aspects of memory and mental health.  
 15 To test our hypothesis, we collected (in aggregate) roughly a century's worth of fitness data.  
 16 We then asked participants to fill out surveys asking them to self-report on different aspects of  
 17 their mental health. We also asked participants to engage in a battery of memory tasks that  
 18 tested their short and long term episodic, semantic, and spatial memory performance. We found  
 19 that participants with similar exercise physical activity habits and fitness profiles tended to also  
 20 exhibit similar mental health and task performance profiles. These effects were task-specific in  
 21 that different exercise physical activity patterns or fitness characteristics varied with different  
 22 aspects of memory, on different tasks. Taken together, these findings provide foundational work  
 23 for designing exercise physical activity interventions that target specific components of cognitive  
 24 performance and mental health by leveraging low-cost fitness tracking devices.

25

## Introduction

26 Engaging in physical activity (exercise) can improve physical fitness by increasing muscle strength  
 27 [9, 20, 23, 34], bone density [1, 8, 21], cardiovascular performance [24, 32], lung capacity [22,  
 28 although see [35]], and endurance [43]. Exercise-Physical activity can also improve mental  
 29 health [2, 4, 10, 25, 31, 33, 40] and cognitive performance [2, 3,  
 30 6, 11].

31 The physical benefits of exercise can be explained by stress-responses of the affected body  
 32 tissues. For example, skeletal muscles that are taxed during exercise exhibit stress responses [28]  
 33 that can in turn affect their growth or atrophy [36]. By comparison, the benefits of exercise

34 physical activity on mental health are less direct. For example, one hypothesis is that ~~exercise~~  
35 physical activity leads to specific physiological changes, such as increased aminergic synaptic  
36 transmission and endorphin release, which in turn act on neurotransmitters in the brain [31].  
37 Speculatively, if different ~~exercise~~physical activity regimens lead to different neurophysiological  
38 responses, one might be able to map out a spectrum of signalling and transduction pathways that  
39 are impacted by a given type, duration, and intensity of ~~exercise~~physical activity in each brain  
40 region. For example, prior work has shown that ~~exercise~~physical activity increases acetylcholine  
41 levels, starting in the vicinity of the exercised muscles [37]. Acetylcholine is thought to play an  
42 important role in memory formation [e.g., by modulating specific synaptic inputs from entorhinal  
43 cortex to the hippocampus, albeit in rodents; 30]. Given the central role that these medial temporal  
44 lobe structures play in memory, changes in acetylcholine might lead to specific changes in memory  
45 formation and retrieval.

46 In the present study, we hypothesize that (a) different ~~exercise regimens~~intensities of physical  
47 activity will have different, quantifiable impacts on cognitive performance and mental health, and  
48 that (b) these impacts will be ~~consistant~~consistent across individuals. To this end, we collected  
49 a year of real-world fitness tracking data from each of 113 participants. We then asked each  
50 participant to fill out a brief survey in which they self-evaluated and self-reported several aspects  
51 of their mental health. Finally, we ran each participant through a battery of memory tasks, which  
52 we used to evaluate their memory performance along several dimensions. We searched the data  
53 for potential associations between memory, mental health, and ~~exercise~~physical activity.

## 54 Methods

55 We ran an online experiment using the Amazon Mechanical Turk (MTurk) platform [13]. We  
56 collected data about each participant's fitness and ~~exercise~~physical activity habits, a variety of  
57 self-reported measures concerning their mental health, and about their performance on a battery  
58 of memory tasks.

59 **Experiment**

60 **Participants**

61 We recruited experimental participants by posting our experiment as a Human Intelligence Task  
62 (HIT) on the MTurk platform. We limited participation to MTurk Workers who had been assigned  
63 a “master worker” designation on the platform, given to workers who score highly across several  
64 metrics on a large number of HITs, according to a proprietary algorithm managed by Amazon. One  
65 ~~criterion~~ embedded into the algorithm is a requirement that master workers must maintain  
66 a HIT acceptance rate of at least 95%. We further limited our participant pool to participants who  
67 self-reported that they were fluent in English and regularly used a Fitbit fitness tracker device.  
68 A total of 160 workers accepted our HIT in order to participate in our experiment. Of these,  
69 we excluded all participants who failed to log into their Fitbit account (giving us access to their  
70 anonymized fitness tracking data), encountered technical issues (e.g., by accessing the HIT using an  
71 incompatible browser, device, or operating system), or who ended their participation prematurely,  
72 before completing the full study. In all, 113 participants contributed usable data to the study.

73 For their participation, workers received a base payment of \$5 per hour (computed in 15  
74 minute increments, rounded up to the nearest 15 minutes), plus an additional performance-based  
75 bonus of up to \$5. Our recruitment procedure and study protocol were approved by Dartmouth’s  
76 Committee for the Protection of Human Subjects. We obtained informed consent using an online  
77 form administered to all prospective participants prior to enrolling them in our study. All methods  
78 were performed in accordance with the relevant guidelines and regulations.

79 **Gender, age, and race.** Of the 113 participants who contributed usable data, 77 reported their  
80 gender as female, 35 as male, and 1 chose not to report their gender. Participants ranged in age  
81 from ~~19–68~~ <sup>19 to 68</sup> years old (25<sup>th</sup> percentile: 28.25 years; 50<sup>th</sup> percentile: 32 years; 75<sup>th</sup> percentile:  
82 38 years). Participants reported their race as White (90 participants), Black or African American (11  
83 participants), Asian (7 participants), Other (4 participants), and American Indian or Alaska Native  
84 (3 participants). One participant opted not to report their race.

85 **Languages.** All participants reported that they were fluent in either 1 and 2 languages (25<sup>th</sup>  
86 percentile: 1; 50<sup>th</sup> percentile: 1; 75<sup>th</sup> percentile: 1), and that they were “familiar” with between 1  
87 and 11 languages (25<sup>th</sup> percentile: 1; 50<sup>th</sup> percentile: 2; 75<sup>th</sup> percentile: 3).

88 **Reported medical conditions and medications.** Participants reported having and/or taking med-  
89 ications pertaining to the following medical conditions: anxiety or depression (4 participants),  
90 recent head injury (2 participants), high blood pressure (1 participant), bipolar disorder (1 partici-  
91 pant), hypothyroidism (1 participant), and other unspecified conditions or medications (1 partici-  
92 pant). Participants reported their current and typical stress levels on a Likert scale as very relaxed  
93 (-2), a little relaxed (-1), neutral (0), a little stressed (1), or very stressed (2). The “current” stress  
94 level reflected participants’ stress at the time they participated in the experiment. Their responses  
95 ranged from -2 to 2 (current stress: 25<sup>th</sup> percentile: -2; 50<sup>th</sup> percentile: -1; 75<sup>th</sup> percentile: 1; typical  
96 stress: 25<sup>th</sup> percentile: 0; 50<sup>th</sup> percentile: 1; 75<sup>th</sup> percentile: 1). Participants also reported their  
97 current level of alertness on a Likert scale as very sluggish (-2), a little sluggish (-1), neutral (0), a  
98 little alert (1), or very alert (2). Their responses ranged from -2 to 2 (25<sup>th</sup> percentile: 0; 50<sup>th</sup> per-  
99 centile: 1; 75<sup>th</sup> percentile: 2). Nearly all (111 out of 113) participants reported that they had normal  
100 color vision, and 15 participants reported uncorrected visual impairments (including dyslexia and  
101 uncorrected near- or far-sightedness).

102 **Residence and level of education.** Participants reported their residence as being located in the  
103 suburbs (36 participants), a large city (30 participants), a small city (23 participants), rural (14 partici-  
104 pants), or a small town (10 participants). Participants reported their level of education as follows:  
105 College graduate (42 participants), Master’s degree (23 participants), Some college (21 partici-  
106 pants), High school graduate (9 participants), Associate’s degree (8 participants), Other graduate  
107 or professional school (5 participants), Some graduate training (3 participants), or Doctorate (2  
108 participants).

109 **Reported water and coffee intake.** Participants reported the number of 8 oz cups of water and  
110 coffee they had consumed prior to accepting the HIT. Water consumption ranged from ~~0-6~~ 0 to 6

111 cups (25<sup>th</sup> percentile: 1; 50<sup>th</sup> percentile: 3; 75<sup>th</sup> percentile: 4). Coffee consumption ranged from ~~0-4~~  
112 0 to 4 cups (25<sup>th</sup> percentile: 0; 50<sup>th</sup> percentile: 1; 75<sup>th</sup> percentile: 2).

113 **Tasks**

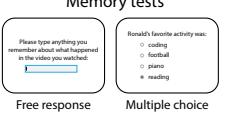
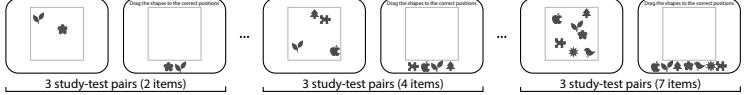
114 Upon accepting the HIT posted on MTurk, each worker was directed to read and fill out a screening  
115 and consent form, and to share access to their anonymized Fitbit data via their Fitbit account. After  
116 consenting to participate in our study and successfully sharing their Fitbit data, participants filled  
117 out a survey and then engaged in a series of memory tasks (Fig. 1). All stimuli and code for running  
118 the full MTurk experiment may be found [here](#).

119 **Survey questions.** We collected the following demographic information from each participant:  
120 their birth year, gender, highest (academic) degree achieved, race, language fluency, and language  
121 familiarity. We also collected information about participants' health and wellness, including about  
122 their vision, alertness, stress, sleep, coffee and water consumption, location of their residence,  
123 activity typically required for their job, and ~~exercise~~ physical activity habits.

124 **Free recall (Fig. 1a).** Participants studied a sequence of four word lists, each comprising 16 words.  
125 After studying each list, participants received an immediate memory test, whereby they were asked  
126 to type (one word at a time) any words they remembered from the just-studied list, in any order.

127 Words were presented for 2 s each, in black text on a white background, followed by a 2 s blank  
128 (white) screen. After the final 2 s pause, participants were given 90 s to type in as many words  
129 as they could remember, in any order. The memory test was constructed such that the participant  
130 could only see the text of the current word they were typing; when they pressed any non-letter  
131 key, the current word was submitted and the text box they were typing in was cleared. This was  
132 intended to prevent participants from retroactively editing their previous responses.

133 The word lists participants studied were drawn from the categorized lists reported by [44]. Each  
134 participant was assigned four unique randomly chosen lists (in a randomized order), selected from  
135 a full set of 16 lists. Each chosen list was then randomly shuffled before presenting the words to

	Main task and immediate memory test				Delayed memory test
a.	1 Free recall	Study words 	Memory test 		5 
b.	2 Naturalistic recall	Watch a short video (The Temple of Knowledge) 	Memory tests 		6 
c.	3 Foreign language flashcards	Study flashcards 	Memory test 		7 
d.	4 Spatial learning	Memorize the positions of increasing numbers of shapes 			N/A

**Figure 1: Battery of memory tasks.** **a. Free recall.** Participants study 16 words (presented one at a time), followed by an immediate memory test where they type each word they remember from the just-studied list. In the delayed memory test, participants type any words they remember studying, from any list. **b. Naturalistic recall.** Participants watch a brief video, followed by two immediate memory tests. The first test asks participants to write out what happened in the video. The second test has participants answer a series of multiple choice questions about the conceptual content of the video. In the delayed memory test, participants (again) write out what happened in the video. **c. Foreign language flashcards.** Participants study a sequence of 10 English-Gaelic word pairs, each presented with an illustration of the given word. During an immediate memory test, participants perform a multiple choice test where they select the Gaelic word that corresponds to the given photograph. During the delayed memory test, participants perform a second multiple choice test, where they select the Gaelic word that corresponds to each of a new set of photographs. **d. Spatial learning.** In each trial, participants study a set of randomly positioned shapes. Next, the shapes' positions are altered, and participants are asked to drag the shapes back to their previous positions. **All panels.** The gray numbers denote the order in which participants experienced each task or test.

<sup>136</sup> the participants. Participants also performed a final delayed memory test where they were given  
<sup>137</sup> 180 s to type out any words they remembered from *any* of the 4 lists they had studied.

<sup>138</sup> Recalled words within an edit distance of 2 (i.e., a Levenshtein Distance less than or equal to  
<sup>139</sup> 2) of any word in the wordpool were “autocorrected” to their nearest match. We also manually  
<sup>140</sup> corrected clear typos or misspellings by hand (e.g., we corrected “hippoptumas” to “hippopota-  
<sup>141</sup> mus”, “zucinni” to “zucchini”, and so on). Finally, we lemmatized each submitted word to match  
<sup>142</sup> the plurality of the matching wordpool word (e.g., “bongo” was corrected to “bongos”, and so  
<sup>143</sup> on). After applying these corrections, any submitted words that matched words presented on the  
<sup>144</sup> just-studied list were tagged as “correct” recalls, and any non-matching words were discarded  
<sup>145</sup> as “errors.” Because participants were not allowed to edit the text they entered, we chose not to  
<sup>146</sup> analyze these putative “errors,” since we could not distinguish typos from true misrememberings.

<sup>147</sup> **Naturalistic recall (Fig. 1b).** Participants watched a 2.5 ~~minute~~ ~minute video clip entitled “The  
<sup>148</sup> Temple of Knowledge.” The video comprises an animated story told to StoryCorps by Ronald  
<sup>149</sup> Clark, who was interviewed by his daughter, Jamilah Clark. The narrator (Ronald) discusses  
<sup>150</sup> growing up living in an apartment over the Washington Heights branch of the New York Public  
<sup>151</sup> Library, where his father worked as a custodian during the 1940s.

<sup>152</sup> After watching the video clip, participants were asked to type out anything they remembered  
<sup>153</sup> about what happened in the video. They typed their responses into a text box, one sentence at a  
<sup>154</sup> time. When the participant pressed the return key or typed any final punctuation mark (“.”, “!”, or  
<sup>155</sup> “?”) the text currently entered into the box was “submitted” and added to their transcript, and the  
<sup>156</sup> text box was cleared to prevent further editing of any already-submitted text. This was intended to  
<sup>157</sup> prevent participants from retroactively editing their previous responses. Participants were given  
<sup>158</sup> up to 10 minutes to enter their responses. After 4 minutes, participants were given the option of  
<sup>159</sup> ending the response period early, e.g., if they felt they had finished entering all ~~of~~ the information  
<sup>160</sup> they remembered. Each participant’s transcript was constructed from their submitted responses by  
<sup>161</sup> combining the sentences into a single document and removing extraneous whitespace characters.  
<sup>162</sup> Following this ~~4–10 minute~~ 4–10-minute free response period, participants were given a series of 10

<sup>163</sup> multiple choice questions about the conceptual content of the story. All participants received the  
<sup>164</sup> same questions, in the same order. Participants also performed a final delayed memory test, where  
<sup>165</sup> they carried out the free response recall task a second time, near the end of the testing session. This  
<sup>166</sup> resulted in a second transcript, for each participant.

<sup>167</sup> **Foreign language flashcards (Fig. 1c).** Participants studied a series of 10 English-Gaelic word  
<sup>168</sup> pairs in a randomized order. We selected the Gaelic language both for its relatively small number  
<sup>169</sup> of native speakers and for its dissimilarity to other commonly spoken languages amongst MTurk  
<sup>170</sup> workers. We verified (via self report) that all of our participants were fluent in English and that  
<sup>171</sup> they were neither fluent nor familiar with Gaelic.

<sup>172</sup> Each word's "flashcard" comprised a cartoon depicting the given word, the English word or  
<sup>173</sup> phrase in lowercase text (e.g., "the boy"), and the Gaelic word or phrase in uppercase text (e.g.,  
<sup>174</sup> "BUACHAILL"). Each flashcard was displayed for 4 s, followed by a 3 s interval (during which  
<sup>175</sup> the screen was cleared) prior to the next flashcard presentation.

<sup>176</sup> After studying all 10 flashcards, participants were given a multiple choice memory test where  
<sup>177</sup> they were shown a series of novel photographs, each depicting one of the 10 words they had  
<sup>178</sup> learned. They were asked to select which (of 4 unique options) Gaelic word went with the given  
<sup>179</sup> picture. The 3 incorrect options were selected at random (with replacement across trials), and the  
<sup>180</sup> order in which the choices appeared to the participant were also randomized. Each of the 10 words  
<sup>181</sup> they had learned were tested exactly once.

<sup>182</sup> Participants also performed a final delayed memory test, where they were given a second set of  
<sup>183</sup> 10 questions (again, one per word they had studied). For this second set of questions participants  
<sup>184</sup> were prompted with a new set of novel photographs, and new randomly chosen incorrect choices  
<sup>185</sup> for each question. Each of the 10 original words they had learned were (again) tested exactly once  
<sup>186</sup> during this final memory test.

<sup>187</sup> **Spatial learning (Fig. 1d).** Participants performed a series of study-test trials where they memo-  
<sup>188</sup> rized the onscreen spatial locations of two or more shapes. During the study phase of each trial,

189 a set of shapes appeared on the screen for 10 s, followed by 2 s of blank (white) screen. During the  
190 test phase of each trial, the same shapes appeared onscreen again, but this time they were vertically  
191 aligned and sorted horizontally in a random order. Participants were instructed to drag (using the  
192 mouse) each shape to its studied position, and then to click a button to indicate that the placements  
193 were complete.

194 In different study-test trials, participants learned the locations of different numbers of shapes  
195 (always drawn from the same pool of 7 unique shapes, where each shape appeared at most one  
196 time per trial). They first performed three trials where they learned the locations of 2 shapes; next  
197 three trials where they learned the locations of 3 shapes; and so on until their last three trials, where  
198 (during each trial) they learned the locations of 7 shapes. All told, each participant performed 18  
199 study-test trials of this spatial learning task (3 trials for each of 2, 3, 4, 5, 6, and 7 shapes).

200 **Fitness tracking using Fitbit devices**

201 To gain access to our study, participants provided us with access to all data associated with their  
202 Fitbit account from the year (365 calendar days) up to and including the day they accepted the HIT.  
203 We filtered out all identifiable information (e.g., participant names, GPS coordinates, etc.) prior to  
204 importing their data.

205 **Collecting and processing Fitbit data**

206 The fitness tracking data associated with participants' Fitbit accounts varied in scope and duration  
207 according to which device the participant owned (Fig. S1), how often the participant wore (and/or  
208 synced) their tracking device, and how long they had owned their device. For example, while all  
209 participants' devices supported basic activity metrics such as daily step counts, only a subset of  
210 the devices with heart rate monitoring capabilities provided information about workout intensity,  
211 resting heart rate, and other related measures. Across all devices, we collected the following infor-  
212 mation: heart rate data, sleep tracking data, logged bodyweight measurements, logged nutrition  
213 measurements, Fitbit account and device settings, and activity metrics.

214 **Heart rate.** If available, we extracted all heart rate data collected by participants' Fitbit device(s)  
215 and associated with their Fitbit profile. Depending on the specific device model(s) and settings, this  
216 included second-by-second, minute-by-minute, daily summary, weekly summary, and/or monthly  
217 summary heart rate information. These summaries include information about participants' aver-  
218 age heart rates, and the amount of time they were estimated to have spent in different "heart rate  
219 zones" (rest, out-of-range, fat burn, cardio, or peak, as defined by their Fitbit profile), as well as an  
220 estimate of the number of estimated calories burned while in each heart rate zone.

221 **Sleep.** If available, we extracted all sleep data collected by participants' Fitbit device(s). Depend-  
222 ing on the specific device model(s) and settings, this included nightly estimates of the duration  
223 and quality of sleep, as well as the amount of time spent in each sleep stage (awake, REM, light, or  
224 deep).

225 **Weight.** If available, we extracted any weight-related information affiliated with participants'  
226 Fitbit accounts within 1 year prior to enrolling in our study. Depending on their specific device  
227 model(s) and settings, this included their weight, body mass index, and/or body fat percentage.

228 **Nutrition.** If available, we extracted any nutrition-related information affiliated with participants'  
229 Fitbit accounts within 1 year prior to enrolling in our study. Depending on their specific account  
230 settings and usage behaviors, this included a log of the specific foods they had eaten (and logged)  
231 over the past year, and the amount of water consumed (and logged) each day.

232 **Account and device settings.** We extracted any settings associated with participants' Fitbit ac-  
233 counts to determine (a) which device(s) and model(s) are associated with their Fitbit account, (b)  
234 time(s) when their device(s) were last synced, and (c) battery level(s).

235 **Activity metrics.** If available, we extracted any activity-related information affiliated with par-  
236 ticipants' Fitbit accounts within 1 year prior to enrolling in our study. Depending on their specific  
237 device model(s) and settings, this included: daily step counts; daily amount of time spent in each

238 activity level (sedentary, lightly active, fairly active, or very active, as defined by their account  
239 settings and preferences); daily number of floors climbed; daily elevation change; and daily total  
240 distance traveled.

241 **Comparing recent versus baseline measurements.**

242 We were interested in separating out potential associations between *absolute* fitness metrics and  
243 *relative* metrics. To this end, in addition to assessing potential raw (absolute) fitness metrics, we  
244 also defined a simple measure of recent changes in those metrics, relative to a baseline:

$$\Delta_{R,B}m = \frac{B \sum_{i=1}^R m(i)}{R \sum_{i=R+1}^{R+B} m(i)},$$

245 where  $m(i)$  is the value of metric  $m$  from  $i - 1$  days prior to testing (e.g.,  $m(1)$  represents the value  
246 of  $m$  on the day the participant accepted the HIT, and  $m(10)$  represents the value of  $m$  9 days prior  
247 to accepting the HIT). We set  $R = 7$  and  $B = 30$ . In other words, to estimate recent changes in any  
248 metric  $m$ , we divided the average value of  $m$  taken over the prior week by the average value of  $m$   
249 taken over the 30 days before that.

250 **Exploratory correlation analyses**

251 We used a bootstrap procedure to identify reliable correlations between different memory-related,  
252 fitness-related, and demographic-related variables. For each of  $N = 10,000$  iterations, we selected  
253 (with replacement) a sample of 113 participants to include. This yielded, for each iteration, a  
254 sampled “data matrix” with one row per sampled participant and one column for each measured  
255 variable. When participants were sampled multiple times in a given iteration, as was often the  
256 case, this matrix contained duplicate rows. Next, we computed the Pearson’s correlation between  
257 each pair of columns. This yielded, for each pair of columns, a distribution of  $N$  bootstrapped  
258 correlation coefficients. If 97.5% or fewer of the coefficients for a given pair of columns had the  
259 same sign, we excluded the pair from further analysis and considered the expected correlation  
260 between those columns to be undefined. If > 97.5% of the coefficients for a given pair of columns

had the same sign (corresponding to a bootstrap-estimated two-tailed  $p$  threshold of 0.05), we computed the expected correlation coefficient as:

$$\mathbb{E}_{i,j}[r] = \tanh\left(\frac{1}{N} \sum_{n=1}^N \tanh^{-1}(\text{corr}(m(i)_n, m(j)_n))\right),$$

where  $m(x)_n$  represents column  $x$  of the bootstrapped data matrix for iteration  $n$ ,  $\tanh$  is the hyperbolic tangent, and  $\tanh^{-1}$  is the inverse hyperbolic tangent. We estimated the corresponding  $p$ -values for these correlations as one minus the proportion of bootstrapped correlations with the same sign, multiplied by two.

## Reverse correlation analyses

We sought to characterize potential associations between the *dynamics* of participants' fitness-related activities leading up to the time they participated in a memory task and their performance on the given task. For each fitness-related variable, we constructed a timeseries matrix whose rows corresponded to timepoints (sampled once per day) leading up to the day the participant accepted the HIT for our study, and whose columns corresponded to different participants. These matrices often contained missing entries, since different participants' Fitbit devices tracked fitness-related activities differently. For example, participants whose Fitbit devices lacked heart rate sensors would have missing entries for any heart rate-related variables. Or, if a given participant neglected to wear their fitness tracker on a particular day, the column corresponding to that participant would have missing entries for that day. To create stable estimates, we smoothed the timeseries of each fitness measure using a sliding window of 1 week. In other words, for each fitness measure, we replaced the "observed value" for each day with the average values of that measure (when available) over the 7-day interval ending on the given day.

In addition to this set of matrices storing timeseries data for each fitness-related variable, we also constructed a memory performance matrix,  $M$ , whose rows corresponded to different memory-related variables, and whose columns corresponded to different participants. For example, one row of the memory performance matrix reflected the average proportion of words (across lists)

285 that each participant remembered during the immediate free recall test, and so on.

286 Given a fitness timeseries matrix,  $F$ , we computed the weighted average and weighted standard  
287 error of the mean of each row of  $F$ , where the weights were given by a particular memory-related  
288 variable (row of  $M$ ). For example, if  $F$  contained participants' daily step counts, we could use  
289 any row of  $M$  to compute a weighted average across any participants who contributed step count  
290 data on each day. Choosing a row of  $M$  that corresponded to participants' performance on the  
291 naturalistic recall task would mean that participants who performed better on the naturalistic recall  
292 task would contribute more to the weighted average timeseries of daily step counts. Specifically,  
293 for each row,  $t$ , of  $F$ , we computed the weighted average (across the  $S$  participants) as:

$$\bar{f}(t) = \sum_{s=1}^S \hat{m}(s)F(t,s),$$

294 where  $\hat{m}$  denotes the normalized min-max scaling of  $m$  (the row of  $M$  corresponding to the chosen  
295 memory-related variable):

$$\hat{m} = \frac{m}{\sum_{s=1}^S \hat{m}(s)},$$

296 where

$$\hat{m} = (1 - \epsilon) \frac{m - \min(m)}{\max(m) - \min(m)} + \epsilon.$$

297 Here,  $\epsilon$  provides a lower bound on the influence of the lowest-weighted participant's data. We  
298 defined  $\epsilon = 0.001$ , ensuring that the lowest-weighted participant had relatively low (but non-zero)  
299 influence. We computed the weighted standard error of the mean as:

$$\text{SEM}_m(f(t)) = \frac{\left| \sum_{s=1}^S (F(t,s) - \bar{f}(t)) \right|}{\sqrt{S}}.$$

300 When a given row of  $F$  was missing data from one or more participants, those participants were  
301 excluded from the weighted average for the corresponding timepoint and the weights (across all  
302 remaining participants) were re-normalized to sum to 1. The above procedure yielded, for each  
303 memory variable, a timeseries of weighted average (and weighted standard error of the mean)

304 fitness tracking values leading up to the day of the experiment.

305 **Results**

306 Before testing our main hypotheses, we examined the behavioral data from each of four memory  
307 tasks: a random word list learning “free recall” task (Fig. 1); a naturalistic recall task whereby par-  
308 ticipants watched a short video and then recounted the narrative; a foreign language “flashcards”  
309 task; and a spatial learning task. Each of the first three tasks (free recall, naturalistic recall, and the  
310 flashcards task) included both an immediate (short term) memory test and a delayed (long term)  
311 memory test. The spatial learning task included only an immediate test. Participants in all four  
312 tasks exhibited general trends and tendencies that have been previously reported in prior work.  
313 We were also interested in characterizing the variability in task performance across participants.  
314 For example, if all participants exhibited near-identical behaviors or performance on a given task,  
315 we would be unable to identify how memory performance on that task varied with mental health  
316 or exercisephysical activity.

317 When participants engaged in free recall of random word lists, they displayed strong primacy  
318 and recency effects [29] on the immediate memory tests (as reflected by improved memory for  
319 early and late list items; Fig. 2a, left and right panels). On the delayed memory test, the recency  
320 effect was substantially diminished (Fig. 3a, left and right panels), consistent with myriad previous  
321 studies [for review see 18]. Participants also tended to cluster their recalls according to the words'  
322 study positions [17] on both the immediate (Fig. 2a, middle panel) and delayed (Fig. 3a, middle  
323 panel) memory tests.

324 When participants engaged in naturalistic recall by recounting the narrative of a short story  
325 video, they reliably and accurately remembered the major narrative events on both the immediate  
326 (Fig. 2b) and delayed (Fig. 3b) tests. This is consistent with prior work showing that memory for  
327 rich narratives is both detailed and accurate [7, 15].

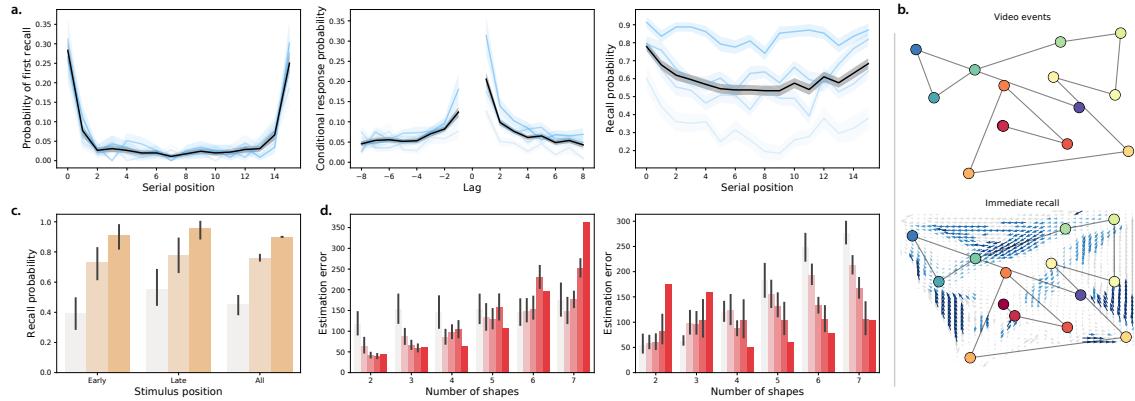
328 Performance on the foreign language flashcards task (immediate: Fig. 2c; delayed: Fig. 3c)  
329 varied substantially across participants, and did not show any clear serial position effects. Partic-

330 participants also displayed substantial variation in performance on the spatial learning task (Fig. 2d).  
331 In general, participants reported the shape's positions more accurately when there were fewer  
332 shapes. However, both the baseline estimation accuracy and the rate of decrease in accuracy as a  
333 function of increasing number of memorized locations varied substantially across participants.

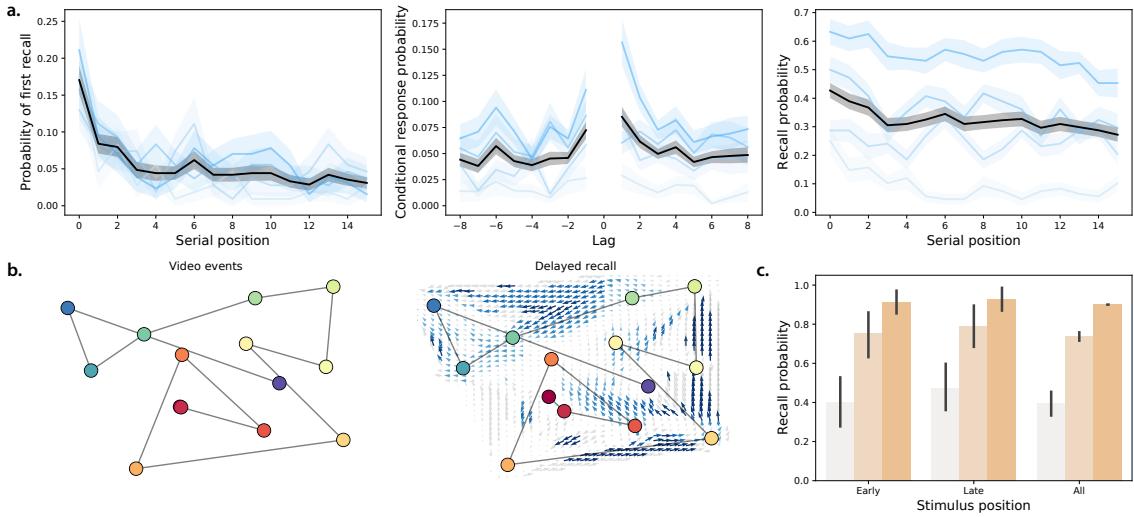
334 In addition to observing substantial across-participant variability in memory performance,  
335 we also observed substantial variability in participants' fitness and activity metrics (Fig. 4). We  
336 examined recent measurements, averaged over the week prior to testing (Fig. 4a), baselined mea-  
337 surements (average over the prior week, divided by the average over the preceding 30 days;  
338 Fig. 4b), along with more gradually varying measures that tended to remain relatively static over  
339 timescales of weeks to months (Fig. 4c). Figure S6 displays across-participant distributions for  
340 a broad selection of these measures, and Figures S7, S8, S9, and S10 show different participants'  
341 fitness metrics, broken down by their performance on different memory tasks.

342 We wondered about potential links between the different aspects of participants' data. For  
343 example, if people who ~~exercised in a particular way~~ engaged in particular intensities of physical  
344 activity also tended to perform better on a given memory task, this could suggest that either (a) some  
345 property intrinsic to participants who exercised in a particular way might also affect their memory  
346 performance on the given task, and/or (b) particular ~~exercise~~ physical activity behaviors could have  
347 a causal impact on memory performance. We carried out an exploratory analysis whereby we used  
348 a bootstrap-based approach to identify reliable correlations between different aspects of memory  
349 performance (Fig. S11), different aspects of fitness (Fig. S12), different demographic attributes  
350 (Fig. S13), and correlations between memory performance, fitness information, and demographic  
351 attributes (Fig. S14). Specifically, we sought to identify correlations that were present in the same  
352 direction (i.e., positive or negative) across different subsets of participants. For each test, we report  
353 the average correlation (taken across 10,000 subsets of participants, chosen with replacement) and  
354 an associated two-tailed  $p$ -value, estimated as

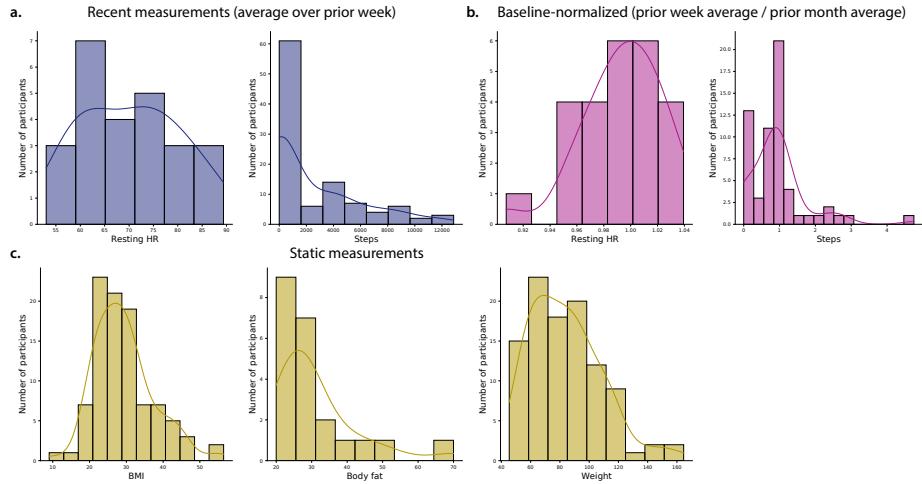
$$p = 2 \times (1 - q),$$



**Figure 2: Immediate memory tests.** **a. Free recall.** Left: probability of recalling each word first as a function of its presentation position. Middle: probability of transitioning between successively recalling the word presented at position  $i$ , followed by word presented at position  $i + \text{Lag}$ . Right: probability of recalling each word as a function of its presentation position. See Figure S2 for additional details. **b. Naturalistic recall.** Top: 2D embedding of a 2.5 min-min video clip; each dot reflects a narrative event (red denotes early events and blue denotes later events). Bottom: 2D embedding of the averaged transcripts of participants' recounts (dots: same format as top panel). The arrows denote the average trajectory directions through the corresponding region of text embedding space, for any participants whose recounts passed through that region. Blue arrows denote statistically reliable agreement across participants ( $p < 0.05$ , corrected). See Figure S3 for additional details. **c. Foreign language flashcards.** Each bar denotes the average proportion of correctly recalled Gaelic-English word pairs from early (first 3), late (last 3), or all (i.e., all 10) study positions. See Figure S4 for additional details. **d. Spatial learning.** Average estimation error in shape locations as a function of the number of shapes. See Figure S5 for additional details. All panels: error bars and error ribbons denote bootstrap-estimated 95% confidence intervals. Shading (saturation) denotes results for different subsets of participants assigned based on their task performance (Figs. S2, S3, S4, and S5 provide information about which performance metrics and values the shading reflects; in general more saturated colors denote participants who performed better on the given task.) In Panel d, participants are grouped in two ways; in the left panel, participants are grouped according to the  $y$ -intercepts of regression lines (estimation error as a function of the number of shapes); in the right panel, participants are grouped according to the slopes of the same regression lines.



**Figure 3: Delayed memory tests.** **a. Free recall.** These panels are in the same format as Figure 2a, but they reflect performance on the delayed free recall task. For additional details see Figure S2. **b. Naturalistic recall.** These panels are in the same format as Figure 2b, but the right panel reflects performance on the delayed naturalistic recall task. For additional details see Figure S3. **c. Foreign language flashcards.** This panel is in the same format as Figure 2c, but it reflects performance on the delayed flashcards test. For additional details see Figure S4.



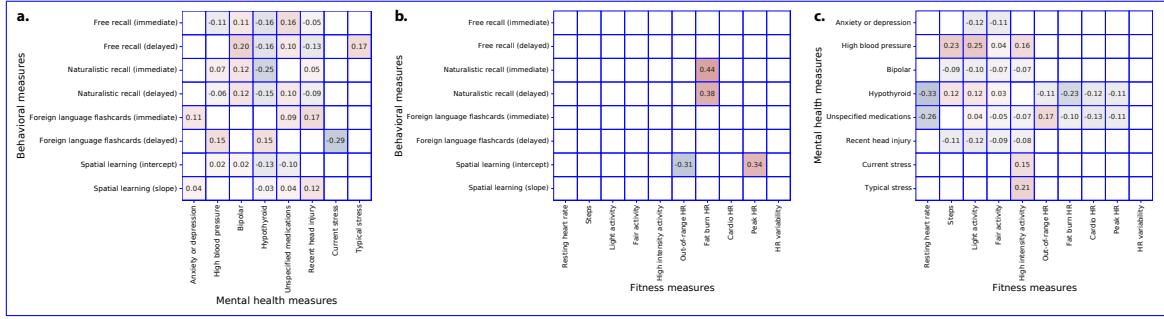
**Figure 4: Fitness measures.** **a. Recent measures.** Resting heart rate (HR) and daily step counts, averaged over the week prior to testing. **Baseline-normalized measures.** Resting heart rate and daily step counts averaged over the week prior to testing, divided by the average resting heart rate and step counts averaged over the preceding month. **Static measures.** Body mass index (BMI), body fat percentage, and weight (in kg). For more information see Figures S6, S7, S8, S9, and S10.

355 where  $q$  is the proportion of those 10,000 subsets that exhibited correlations in the same direction  
356 (see *Exploratory correlation analyses*). When all 10,000 randomly chosen subsets of participants ex-  
357 hibited correlations in the same direction (i.e., all positive correlations or all negative correlations),  
358 we report the  $p$ -value as  $p < 0.0001$ .

359 Several patterns emerged from these analyses. First, we found that participants' performance  
360 on the (within-task) immediate versus delayed memory tests from the free recall, naturalistic  
361 recall, and foreign language flashcards tasks were positively correlated ( $rs > 0.25, ps < 0.003$ ).  
362 This suggests that, within each of these tasks, similar processes or constraints may influence both  
363 short term and long term information retrieval. We also found reliable across-task correlations  
364 between participants' (immediate and delayed) performance on the free recall and foreign language  
365 flashcards tasks ( $rs > 0.3, ps < 0.03$ ).

366 A large number of fitness-related measures displayed reliable correlations (for a complete re-  
367 port, see Fig. S12). For example, body mass index (BMI) and weight were correlated ( $r = 0.91, p <$   
368  $0.0001$ ). Resting heart rate over the prior week was negatively correlated with recent low-to-  
369 moderate-intensity ("fat burn") cardiovascular activity levels ( $r = 0.70, p = 0.0004$ ). Participants'  
370 peak heart rates (averaged over the prior week) were also negatively correlated with recent in-  
371 creases in step counts and daily elevation gains ( $rs < -0.26, ps < 0.03$ ), where recent changes  
372 were defined as the average values over the seven days leading up to the test day divided by  
373 the average values over the preceding 30 days. Several demographic attributes (Fig. S13) dis-  
374 played trivial correlations (e.g., participants identifying as male never reported identifying as  
375 female, and so on). We also observed a negative correlation between reported stress and alertness  
376 ( $r = -0.44, p < 0.0001$ ), and positive correlations between the reported clarity of the instructions  
377 for all tasks ( $rs > 0.26, ps < 0.02$ ).

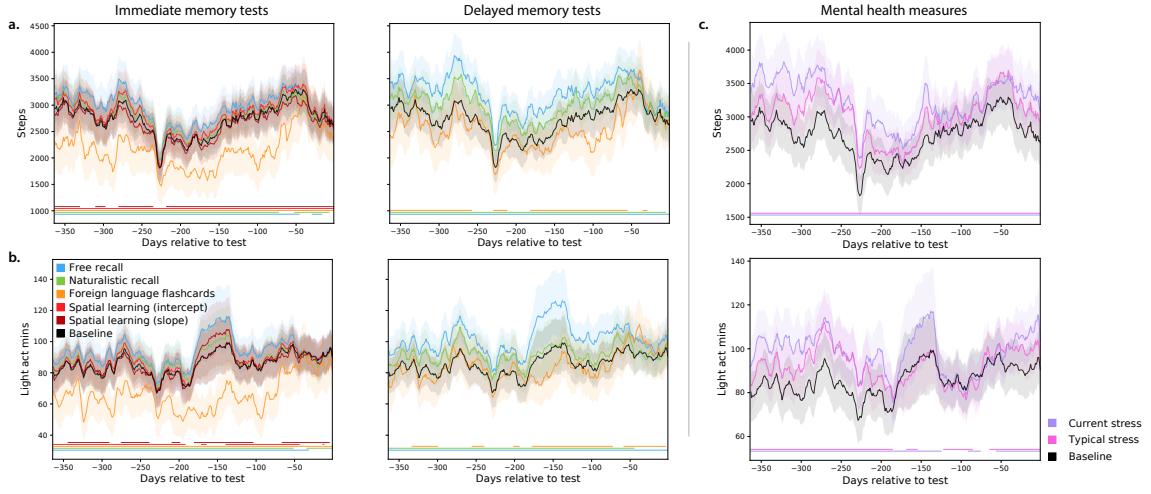
378 We also found reliable correlations between participants' fitness and demographic measures  
379 and their behaviors in different tasks (for a complete report, see Fig. S14). For example, recent  
380 low-to-moderate-intensity ("fat burn") cardiovascular activity was positively correlated with im-  
381 mediate ( $r = 0.38, p = 0.03$ ) and delayed ( $r = 0.38, p = 0.029$ ) recall performance on the naturalistic  
382 memory task. Recent increases in moderate-intensity ("cardio") activity over the prior 7 days



**Figure 5: Memory performance differs according to mental health measures. Summaries of correlations between behavioral, fitness, and mental health measures.** The reported values in the [table](#)s reflect correlations between each behavioral measure and mental health measure. Only statistically reliable correlations ( $p < 0.05$ , corrected) are displayed. **a. Correlations between behavioral and mental health measures.** We adjusted each task's behavioral measure(s) such that more positive values reflect better performance on the given task. We used participants' mean recall accuracy to characterize performance on the free recall and foreign language flashcards tasks, and mean precision to characterize performance on the naturalistic recall tasks. We characterized performance on the spatial learning task using the (inverted and normalized) intercepts and slopes of linear regressions on mean estimation errors as a function of the number of studied shapes (also see Figs. 2, 3, S2, S3, S4, and S5). **For each mental health measure, more positive values denote greater severity of the given measure.** Typical and current stress levels were measured by self report. Mental health information was inferred using each participants' list of self-reported medications (see *Methods*). Positive correlations indicate that better performance on a given behavioral task is associated with more severe mental health phenotypes. **b. Correlations between fitness and mental health measures.** For each fitness measure, more positive values denote higher observed scores (i.e., higher resting heart rate, more minutes of activity or time spent in each heart rate zone, or greater heart rate variability). The mental health measures in this panel were treated as in Panel a. **c. Correlations between fitness and behavioral measures.** Each measure reflected in this panel was treated as in Panels a and b.

(relative to the preceding 30 days) was also positively correlated with immediate naturalistic recall performance ( $r = 0.48, p = 0.003$ ) and immediate recall performance on the foreign language flashcards task ( $r = 0.43, p = 0.048$ ). Recent high-intensity (“peak”) activity was positively correlated with performance on the spatial learning task ( $r = 0.34, p < 0.0001$ ), as were recent increases in high-intensity activity (prior 7 days versus preceding 30 days;  $r = 0.41, p = 0.01$ ). Mental health indicators, such as self-reported stress levels and medications were also associated with differences in memory (Figs. ??5, S14). For example, self-reported stress levels at the time of test were negatively correlated with performance on the delayed memory test for the foreign language flashcards task ( $r = -0.29, p = 0.038$ ), whereas participants who were medicated for anxiety and depression tended to perform slightly (but reliably) *better* on the immediate memory test for the foreign language flashcards task ( $r = 0.11, p < 0.0001$ ).

The above analyses indicate that recent differences in fitness-related activity are associated with differences in memory performance and mental health measures. Although the analyses treated these measures on average or in aggregate, many of the measures we collected are dynamic. For example, the amount or intensity of physical *exercise activity* people engage in can vary over time, and so on. We wondered whether the dynamics of fitness-related measures might relate to memory performance and/or mental health measures. To this end, we carried out a series of reverse correlation analyses (see *Reverse correlation analyses*) to examine whether participants with different cognitive or mental health profiles also tended to display differences in fitness-related measures over time. In particular, we examined fitness data collected from participants’ Fitbit devices over the year prior to their test day in our study. Several example findings are summarized in Figure 6. We found that participants who performed well on the immediate and delayed free recall memory tests and on the naturalistic recall tests tended to be more active than participants who performed poorly on those tests (Figs. 6a, b; S15). Conversely, participants who performed well on the immediate and delayed foreign language flashcards tasks tended to be *less* active. These differences were present even a full year before the testing day. We also found substantial variability across people with different (self-reported) mental health profiles (Figs. 6c, S18). Due to small sample sizes of individuals exhibiting several mental health dimensions, it is difficult to distinguish generalizable



**Figure 6: Dynamics of physical activity varies with memory performance and mental health measures.** **a. Daily step counts.** Each timecourse is weighted by either performance on immediate recall tests (left panel) or on delayed recall tests (right panel). The black (baseline) timecourses display the (unweighted) average across all participants. **b. Daily duration (in minutes) of low-intensity physical activity.** Timecourses are displayed in the same format and color scheme as those in Panel A. Analogous timecourses for additional fitness-related measures may be found in Figures S15, S16, and S17. **c. Timecourses of physical activity, weighted by mental health measures.** The timecourses in each panel display the average daily step counts (top panel) or duration of low-intensity activity (bottom panel). The colored lines show average activity dynamics weighted by self-reported stress levels at the start of the experiment (purple) and self-reported “typical” stress levels (pink). The baseline curves (black) display the average across all participants (re-plotted in Panel C to illustrate scale differences across panels). Timecourses for additional mental health-related and fitness-related measures may be found in Figures S18, S19, and S20. Error ribbons in all panels denote the standard error of the mean. Horizontal lines below each panel’s timecourses denote intervals over which each weighted measure (color) differs from the unweighted baseline (via a paired sample two-sided  $t$ -test of the weighted mean values for each measure within a 30-day-30-day window around each timepoint; horizontal lines denote  $p < 0.05$ , corrected).

411 trends from individual differences that one or two individuals happened to exhibit. However,  
412 several large-sample-size trends emerged. For example, participants who reported higher levels  
413 of stress also tended to be slightly more physically active than participants who reported lower  
414 stress levels. We found analogous differences in other activity-related measures (Figs. S15 and S18),  
415 cardiovascular measures (Figs. S16 and S19), and sleep-related measures (Figs. S17 and S20). Taken  
416 together, the analyses suggest that cognitive and mental health differences are also associated with  
417 differences in the dynamics of physical health measures.

## 418 Discussion

419 After collecting a year's worth of fitness-tracking data from each of 113 participants, we ran  
420 each participant in a battery of memory tasks and had them fill out a series of demographic  
421 and mental health-related questions. We found that the associations between fitness-related ac-  
422 tivities, memory performance, and mental health ~~were heterogeneous. Our results suggest that~~  
423 ~~engagement in particular physical activities (e.g., differing in time relative to the test day, intensity,~~  
424 ~~duration, etc.) is also reflected in participants' memory performance patterns across tasks and in~~  
425 ~~participants' mental health attributes are complex. For example, participants who tended to engage~~  
426 ~~in a particular intensity of physical activity also tended to perform better on some memory tasks~~  
427 ~~but worse on others. This suggests that engaging in one form or intensity of physical activity will~~  
428 ~~not necessarily affect all aspects of cognitive or mental health equally (or in the same direction).~~

429 A number of prior studies have shown that engaging in exercise can improve cognitive and  
430 mental health [2, 3, 4, 6, 10, 11, 12, 14, 25, 26, 27, 31, 33, 40]. The majority of these studies ask  
431 participants in an "exercise intervention" condition (where participants engage in a designated  
432 physical activity or training regimen) or a "control" condition (where participants do not engage in  
433 the designated activity or training) to perform cognitive tasks or undergo mental health screening.  
434 In other words, most primary studies treat "physical activity" as a binary variable that either is  
435 or is not present for each participant. Most prior studies also track or manipulate exercise over  
436 relatively short durations (typically on the order of days or weeks). Our current work indicates

437 that the true relations between physical activity, cognitive performance, and mental health may  
438 be non-monotonic and heterogeneous across activities, tasks, and mental health measures. These  
439 relations can also unfold over much longer timescales than have been previously identified (on the  
440 order of months; Fig. 6). However, despite the complexities of the structures of these associations,  
441 we also found that they were often remarkably consistent across people. For example, as displayed  
442 in Figures 5 and S14, many of the associations between fitness, behavioral, and mental health  
443 measures were consistent across over 97.5% of 10,000 randomly chosen subsets of participants.

444 One important limitation of our study is that we cannot distinguish correlations between  
445 different measures from potential causal effects. For example, we cannot know (from our study)  
446 whether engaging in particular forms of *exercise*-physical activity causes changes in memory  
447 performance or mental health, or whether (alternatively) people who tend to engage in similar  
448 forms of *exercise*-physical activity also happen to exhibit similar memory and/or mental health  
449 profiles. In other words, an overlapping set of processes or person-specific attributes may lead  
450 someone to both form particular habits around *exercise*-physical activity and display high or low  
451 performance on a given memory test. We do not know whether memory performance or aspects  
452 of mental health might be manipulated or influenced by changing the patterns of physical activity  
453 someone engages in. For this reason, we have been careful to frame our findings as correlations  
454 and associations, rather than to imply knowledge about a causal direction of our findings.

455 Although the present study cannot reveal causal effects, a large prior literature provides some  
456 insight into potential causal effects by examining the neural and cognitive effects of a variety of  
457 exercise interventions [5, 16, 19, 38, 39, 41, 42]. A limitation of that prior work is that most of these  
458 studies examine how relatively short-term changes in *exercise*-physical activity (e.g., on timescales  
459 of hours to days or, rarely, weeks to months) affect a cognitive performance on single task or aspect  
460 of mental health. The present study examines longer-term *exercise*-physical activity (over a full  
461 year), and relates long-term *exercise*-physical activity history to performance on a variety of tasks  
462 and to a variety of mental health dimensions.

463 To the extent that *exercise*-physical activity does provide a non-invasive means of manipulating  
464 cognitive performance and mental health, our work may have exciting implications for cognitive

465 enhancement. For example, one might imagine building a recommendation system that suggests a  
466 particular **exercise****physical activity** regimen tailored to improve a specific aspect of an individual's  
467 cognitive performance (e.g., the efficacy of a student's study session for an upcoming exam) or  
468 mental health (e.g., reducing symptoms of anxiety before an important meeting). Just as strength  
469 training may be customized to target a specific muscle group, or to improve performance on a  
470 specific physical task, similar principles might also be applied to target specific improvements in  
471 cognitive fitness and mental health.

## 472 **Acknowledgements**

473 We acknowledge useful discussions with David Bucci, Emily Glasser, Andrew Heusser, Abigail  
474 Bartolome, Lorie Loeb, Lucy Owen, and Kirsten Ziman. Our work was supported in part by the  
475 Dartmouth Young Minds and Brains initiative, and by NSF grant number 2145172 to J.R.M. The  
476 content is solely the responsibility of the authors and does not necessarily represent the official  
477 views of our supporting organizations. This paper is dedicated to the memory of David Bucci, who  
478 helped to inspire the theoretical foundations of this work. Dave served as a mentor and colleague  
479 on the project prior to his passing.

## 480 **Data and code availability**

481 All analysis code and data used in the present manuscript may be found [here](#).

## 482 **Author contributions**

483 Concept: J.R.M. and G.M.N. Experiment implementation and data collection: G.M.N. Analyses:  
484 J.R.M., G.M.N., E.C., and P.C.F. Writing: J.R.M. with input from all authors.

485 **Competing interests**

486 The authors declare no competing interests.

487 **References**

- 488 [1] Bassey, E. J. and Ramsdale, S. J. (1994). Increase in femoral bone density in young women  
489 following high-impact exercise. *Osteoporosis International*, 4:72–75.
- 490 [2] Basso, J. C. and Suzuki, W. A. (2017). The effects of acute exercise on mood, cognition, neuro-  
491 physiology, and neurochemical pathways: a review. *Brain Plasticity*, 2(2):127–152.
- 492 [3] Brisswalter, J., Collardeau, M., and René, A. (2002). Effects of acute physical exercise charac-  
493 teristics on cognitive performance. *Sports Medicine*, 32:555–566.
- 494 [4] Callaghan, P. (2004). Exercise: a neglected intervention in mental health care? *Psychiatric and*  
495 *Mental Health Nursing*, 11(4):476–483.
- 496 [5] Chang, Y.-K., Chu, C.-H., Wang, C.-C., Wang, Y.-C., Song, T.-F., Tsai, C.-L., and Etnier, J. L.  
497 (2015). Dose-response relation between exercise duration and cognition. *Medicine and Science in*  
498 *Sports and Exercise*, 47:159–165.
- 499 [6] Chang, Y. K., Labban, J. D., Gapin, J. I., and Etnier, J. L. (2012). The effects of acute exercise on  
500 cognitive performance: a meta-analysis. *Brain Research*, 1453:87–101.
- 501 [7] Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., and Hasson, U. (2017). Shared  
502 memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*,  
503 20(1):115.
- 504 [8] Chilibeck, P. D., Sale, D. G., and Webber, C. E. (2012). Exercise and bone mineral density. *Sports*  
505 *Medicine*, 19:103–122.

- 506 [9] Crane, J. D., MacNeil, L. G., and Tarnopolsky, M. A. (2013). Long-term aerobic exercise is  
507 associated with greater muscle strength throughout the life span. *The Journals of Gerontology: Series A*, 68(6):631–638.
- 509 [10] Deslandes, A., Moraes, H., Ferreira, C., Veiga, H., Silveira, H., Mouta, R., Pompeu, F. A. M. S.,  
510 Coutinho, E. S. F., and Laks, J. (2009). Exercise and mental health: many reasons to move.  
511 *Neuropsychobiology*, 59:191–198.
- 512 [11] Etnier, J. L., Nowell, P. M., Landers, D. M., and Sibley, B. A. (2006). A meta-regression to  
513 examine the relationship between aerobic fitness and cognitive performance. *Brain Research: Brain Research Reviews*, 52(1):119–130.
- 515 [12] Gordon, B. R., McDowell, C. P., Lyons, M., and Herring, M. P. (2017). The effects of resistance  
516 exercise training on anxiety: a meta-analysis and meta-regression analysis of randomized  
517 controlled trials. *Sports Medicine*, 47(12):2521–2532.
- 518 [13] Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., Halpern,  
519 D., Hamrick, J. B., and Chan, P. (2016). psiTurk: an open-source framework for conducting  
520 replicable behavioral experiments online. *Behavior Research Methods*, 48(3):829–842.
- 521 [14] Herring, M. P., O'Connor, P. J., and Dishman, R. K. (2010). The effects of exercise training on  
522 anxiety symptoms among patients: a systematic review. *Archives of Internal Medicine*, 170(4):321–  
523 331.
- 524 [15] Heusser, A. C., Fitzpatrick, P. C., and Manning, J. R. (2021). Geometric models reveal behavioral  
525 and neural signatures of transforming naturalistic experiences into episodic memories.  
526 *Nature Human Behavior*, 5:905–919.
- 527 [16] Imboden, C., Gerber, M., Beck, J., Eckert, A., Pühse, U., Holsboer-Trachsler, E., and Hatzinger,  
528 M. (2019). Effects of aerobic exercise as add-on treatment for inpatients with moderate to severe  
529 depression on depression severity, sleep, cognition, psychological well-being, and biomarkers:  
530 study protocol, description of study population, and manipulation check. *Frontiers in Psychiatry*,  
531 10(262):doi.org/10.3389/fpsyg.2019.00262.

- 532 [17] Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory and Cognition*,  
533 24:103–109.
- 534 [18] Kahana, M. J. (2012). *Foundations of human memory*. Oxford University Press, New York, NY.
- 535 [19] Kamijo, K., Nishihira, Y., Higashiura, T., and Kuroiwa, K. (2007). The interactive effect of  
536 exercise intensity and task difficulty on human cognitive processing. *International Journal of  
537 Psychophysiology*, 65(2):114–121.
- 538 [20] Knuttgen, H. G. (2007). Strength training and aerobic exercise: comparison and contrast.  
539 *Journal of Strength and Conditioning Research*, 21(3):973–978.
- 540 [21] Layne, J. E. and Nelson, M. E. (1999). The effects of progressive resistance training on bone  
541 density: a review. *Medicine and Science in Sports and Exercise*, 31(1):25–30.
- 542 [22] Lazovic-Popovic, B., Zlatkovic-Svenda, M., Durmic, T., Djelic, M., Saranovic, D., and Zugic, V.  
543 (2016). Superior lung capacity in swimmers: some questions, more answers! *Revista Portuguesa  
544 de Pneumologia*, 22(3):151–156.
- 545 [23] Lindh, M. (1979). Increase of muscle strength from isometric quadriceps exercises at different  
546 knee angles. *Scandinavian Journal of Rehabilitation Medicine*, 11(1):33–36.
- 547 [24] Maiorana, A., O'Driscoll, G., Cheetham, C., Collis, J., Goodman, C., Rankin, S., Taylor, R.,  
548 and Green, D. (2000). Combined aerobic and resistance exercise training improves functional  
549 capacity and strength in CHF. *Journal of Applied Physiology*, 88(1565–1570).
- 550 [25] Mikkelsen, K., Stojanovska, L., Polenakovic, M., Bosevski, M., and Apostolopoulos, V. (2017).  
551 Exercise and mental health. *Maturitas*, 106:48–56.
- 552 [26] Morres, I. D., Hatzigeorgiadis, A., Stathi, A., Comoutos, N., Arpin-Cribbie, C., Krommidas, C.,  
553 and Theodorakis, Y. (2018). Aerobic exercise for adult patients with major depressive disorder in  
554 mental health services: a systematic review and meta-analysis. *Depression and Anxiety*, 36(1):39–  
555 53.

- 556 [27] Morres, I. D., Tzouma, N.-A., Hatzigeorgiadis, A., Krommidas, C., Kotronis, K. V., Dafopoulos,  
557 K., Theodorakis, Y., and Comoutos, N. (2022). Exercise for perinatal depressive symptoms: a  
558 systematic review and meta-analysis of randomized controlled trials in perinatal health services.  
559 *Journal of Affective Disorders*, 298:26–42.
- 560 [28] Morton, J. P., Kayani, A. C., McArdle, A., and Drust, B. (2009). The exercise-induced stress  
561 response of skeletal muscle, with specific emphasis on humans. *Sports Medicine*, 39:643–662.
- 562 [29] Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology:*  
563 *General*, 64:482–488.
- 564 [30] Palacios-Filardo, J., Udkakis, M., Brown, G. A., Tehan, B. G., Congreve, M. S., Nathan, P. J.,  
565 Brown, A. J. H., and Mellor, J. R. (2021). Acetylcholine prioritises direct synaptic inputs from  
566 entorhinal cortex to CA1 by differential modulation of feedforward inhibitory circuits. *Nature  
567 Communications*, 12(5475):doi.org/10.1038/s41467-021-25280-5.
- 568 [31] Paluska, S. A. and Schwenk, T. L. (2000). Physical activity and mental health. *Sports Medicine*,  
569 29(3):167–180.
- 570 [32] Pollock, M. L., Franklin, B. A., Balady, G. J., Chaltman, B. L., Fleg, J. L., Fletcher, B., Limacher,  
571 M., na, I. L. P., Stein, R. A., Williams, M., and Bazzarre, T. (2000). Resistance exercise in individuals  
572 with and without cardiovascular disease. *Circulation*, 101:828–833.
- 573 [33] Raglin, J. S. (1990). Exercise and mental health. *Sports Medicine*, 9:323–329.
- 574 [34] Rogers, M. A. and Evans, W. J. (1993). Changes in skeletal muscle with aging: effects of  
575 exercise training. *Exercise and Sport Sciences Reviews*, 21:65–102.
- 576 [35] Roman, M. A., Rossiter, H. B., and Casaburi, R. (2016). Exercise, ageing and the lung. *European  
577 Respiratory Journal*, 48:1471–1486.
- 578 [36] Schiaffino, S., Dyar, K. A., Cicliot, S., Blaauw, B., and Sandri, M. (2013). Mechanisms regulating  
579 skeletal muscle growth and atrophy. *The febs Journal*, 280(17):4294–4314.

- 580 [37] Shoemaker, J. K., Halliwill, J. R., Hughson, R. L., and Joyner, M. J. (1997). Contributions of  
581 acetylcholine and nitric oxide to forearm blood flow at exercise onset and recovery. *Vascular*  
582 *Physiology*, 273(5):2388–2395.
- 583 [38] Sinha, N., Berg, C. N., Yassa, M. A., and Gluck, M. A. (2021). Increased dynamic flexibility in  
584 the medial temporal lobe network following an exercise intervention mediates generalization  
585 of prior learning. *Neurobiology of Learning and Memory*, 177:107340.
- 586 [39] Suwabe, K., Hyodo, K., Byun, K., Ochi, G., Yassa, M. A., and Soya, H. (2017). Acute moderate  
587 exercise improves mnemonic discrimination in young adults. *Hippocampus*, 27(3):229–234.
- 588 [40] Taylor, C. B., Sallis, J. F., and Needle, R. (1985). The relation of physical activity and exercise  
589 to mental health. *Public Health Reports*, 100(2):195–202.
- 590 [41] Tomporowski, P. D. (2003). Effects of acute bouts of exercise on cognition. *Acta Psychologica*,  
591 112(3):297–324.
- 592 [42] Vidoni, E. D., Johnson, D. K., Morris, J. K., Van Sciver, A., Greer, C. S., Billinger, S. A.,  
593 Donnelly, J. E., and Burns, J. M. (2015). Dose-response of aerobic exercise on cognition: A  
594 community-based, pilot randomized controlled trial. *PLoS One*, 10(7):1–13.
- 595 [43] Wilmore, J. H. and Knuttgen, H. G. (2003). Aerobic exercise and endurance. *The Physician and*  
596 *Sportsmedicine*, 31(5):45–51.
- 597 [44] Ziman, K., Heusser, A. C., Fitzpatrick, P. C., Field, C. E., and Manning, J. R. (2018). Is  
598 automatic speech-to-text transcription ready for use in psychological experiments? *Behavior*  
599 *Research Methods*, 50:2597–2605.